



Københavns Universitet

Comparing linear probability model coefficients across groups

Holm, Anders; Ejrnæs, Mette; Karlson, Kristian Bernt

Published in:
Quality and Quantity

DOI:
[10.1007/s11135-014-0057-0](https://doi.org/10.1007/s11135-014-0057-0)

Publication date:
2015

Document Version
Peer reviewed version

Citation for published version (APA):
Holm, A., Ejrnæs, M., & Karlson, K. B. (2015). Comparing linear probability model coefficients across groups. *Quality and Quantity*, 49(5), 1823-1834. <https://doi.org/10.1007/s11135-014-0057-0>

Comparing linear probability model coefficients across groups

Ander Holm^{a,c}

Mette Ejrnæs^b

Kristian Karlson^a

4,848 words (including all text, references and notes)

6 tables

2 appendices

**THIS PAPER IS PUBLISHED ONLINE IN
QUALITY & QUANTITY, 2014, JULY 26**

Link: <http://link.springer.com/article/10.1007/s11135-014-0057-0>

This is a post-print (i.e. final draft post-refereeing) version according to SHERPA/ROMEO

a: Department of Sociology, University of Copenhagen, b: Department of Economics, University of Copenhagen, c: The Danish National Centre for social Research.

Corresponding author: Anders Holm, Department of Sociology, University of Copenhagen, Øster Farimagsgade 5, bld. 16, DK-1014 Copenhagen K, E-mail: ah@soc.ku.dk, Phone: +45 40935698.

Comparing linear probability model coefficients across groups

Abstract

This article offers a formal identification analysis of the problem in comparing coefficients from linear probability models between groups. We show that differences in coefficients from these models can result not only from genuine differences in effects, but also from differences in one or more of the following three components: outcome truncation, scale parameters and distributional shape of the predictor variable. These results point to limitations in using linear probability model coefficients for group comparisons. We also provide Monte Carlo simulations and real examples to illustrate these limitations, and we suggest a restricted approach to using linear probability model coefficients in-group comparisons.

Keywords: Group differences, Interaction terms, linear probability model, Logit, Probit

Comparing linear probability model coefficients across groups

1. Introduction

Empirical research in social science often involves comparisons of effects between groups. For example, researchers might be interested in the extent to which the gender gap in political attitudes varies across race, or whether the returns to education have changed over cohorts born in the 20th century. However, although differences in effects between groups are straightforward to estimate and interpret in linear models, a growing literature in quantitative social science methodology shows that the principles applying to linear models do not readily extend to nonlinear models for discrete outcomes, such as the logit or probit (Ai and Norton 2003; Allison 1999; Athey and Imbens 2006; Greene 2010; Karaca-Mandic et al. 2011; Long 2009; Mare 2006; Norton et al. 2004; Swait and Louviere 1993; Williams 2009; Xie 2011). As this class of models is widely used in applied social research, this situation is unfortunate and warrants an acceptable solution.

In response to these limitations, scholars have looked to various alternatives. Perhaps most prominently figures the linear probability model (LPM), which has resurfaced as an attractive solution to the group comparison issue. The LPM consistently estimates the conditional expectation of the outcome and has a straightforward interpretation (Greene 2011). In economics, the linear probability model is now widely used. A search identified 19 articles published 2007 through 2011 in the *Quarterly Journal of Economics* using the LPM, amounting to nine percent of all published articles in that period. Among those 19 articles 11 compared coefficients between groups.¹ However, sociologists appear still to prefer the logistic response model. A search of articles published in the same period in the *American*

¹ We notice that probit model is also widely used in economics. In our search, we identified 24 articles in QJE using the probit, with the vast majority (23) reporting marginal effects implied by the probit. Of the 24 articles, we identified eight articles making some form of comparison between groups (reporting marginal effects).

Sociological Review returned one article using the LPM, while we identified 60 articles using logit models, accounting for one-third of all articles published in that period. Of those 60 articles 32 reported some form of group differences in effects.

Despite these apparent differences in practice between economics and sociology, modelling discrete outcome variables continues to be central to social science research. However, while we now much better understand the problem in comparing logit or probit coefficients between groups² or evaluating interaction terms on the margin in these models, little is known about the interpretation of group differences in LPM coefficients.³ In this article we offer a formal identification analysis of this issue. We depart from the assumption that the binary outcome of interest is generated by a linear latent variable model (LLM), as is often assumed in the derivation of nonlinear probability models (e.g. Amemiya 1981). We then derive the relationship between coefficients in the LLM and the LPM, and examine the extent to which LPM can identify group differences in “true” effects.

Our study shows that group differences in LPM coefficients – over and above those that result from differences in “true” effects – can result from group differences in one or more of the following three components: (1) truncation of the outcome, (2) scale parameters and (3) shape of the distribution of the predictor variable. To illustrate how these issues can play out we present some simple examples using data from the U.S. General Social Survey. We conclude the article by presenting a restricted approach to comparing LPM coefficients across groups.

² Allison (1999) and Williams (2009) suggest the heteroskedastic (location-scale) probit or logit model as a possible solution to these issues. However, the utility of these models is disputed. Long (2009) and Breen et al. (Forthcoming) show that these models cannot separate coefficient differences between groups from differences in latent error dispersion, unless very realistic and thus unrealistic assumptions are maintained. We therefore refrain from considering this class of models in this paper.

³ Exceptions are Lewbel et al. (2012) and Athey and Imbens (2009) who show that the LPM does not identify the treatment effect of interest in difference in differences (DiD) models. As DiD models involves comparing trends among groups, their results can be taken to support the results we report in this article.

2. Interpretation of Linear probability models and alternative solutions

In this section, we give some illustrative examples from recent research of interpretations of LPM coefficients in terms of an underlying linear regression model with comparison across groups. In other words, these are examples in which researchers want to infer mechanisms generated by an underlying linear model but where only a dichotomous version of the outcome variable is observed.

Anwar et al (2012) use, among other outcomes, a dependent variable indicating whether a defendant was convicted of at least one charged crime. They interpret effects of regression variables from a LPM as effects on “conviction rates”. Further, they use interaction terms between defendant race (black) and whether there is a black in the jury pool). Also, Fisman et al. (2006) use binary dating decisions as indicators of potential partner’s attractiveness in a dating experiment. They estimate LPM’s on dating decisions using various potential partner characteristics to infer which and how men and women value potential partner’s attractiveness. They further use interaction terms between different partner characteristics. Finally, Meghir and Palme (2005) use binary indicators for the final completed education as indicator of “years of education” and “level of education”. They use a difference in differences approach to study the effect of educational reforms on educational outcomes. The difference in difference approach essentially amounts comparison across groups.

In sum, in these examples, researchers use the LPM with the aim of inferring effects from underlying linear regression coefficients and often, as these examples demonstrate, researcher’s compares LPM regression coefficients between groups.

3. Relationship between coefficients

To show the relationship between coefficients from the LPM and the LLM, assume that a continuous latent variable, y^* , is a linear function of a continuous predictor variable, x :

$$y^* = \alpha + \beta^* x + u \quad (1)$$

where u is an i.i.d. error term assumed to be independent of x with mean zero and standard deviation σ . The model in (1) is what we refer to as the latent linear model (LLM). The true parameter, β^* , can be expressed in terms of covariances:

$$\beta^* = \frac{\text{cov}(x, y^*)}{\text{var}(x)} \quad (2)$$

The OLS estimator will converge to β^* and, in what follows, we refer to population level parameters, not their finite sample counterparts (e.g., $\beta^* = \text{plim}_{n \rightarrow \infty} \hat{\beta}_{ols}^*$).

Now assume that we observe a binary realisation of y^* , y , such that

$$\begin{aligned} y &= 1 && \text{if } y^* > 0 \\ y &= 0 && \text{otherwise.} \end{aligned} \quad (3)$$

The parameters in the model defined by (1) and (3) determine the distribution of the observed binary y . Under the assumption that this LLM has generated the data we ask what the corresponding LPM coefficient of x identifies in terms of the LLM coefficient of x . Because the distribution of the error term, u , is unknown we do so without placing any further assumptions on u . Therefore, we let the standardised distribution of u/σ be characterised by the p.d.f. f and the c.d.f. F .

Write the LPM as

$$y = 1(\mu + \beta^* x + v) \quad (4)$$

where v is a random error term⁴ with $E(v) = 0$ and

$$\beta^{LPM} = \frac{\text{cov}(x, y)}{\text{var}(x)} \quad (5)$$

which gives the conditional expectation of y given x , $E(Y | x) = \beta^* x$, (Greene 2011).

⁴ The error term is inherently heteroscedastic (Goldberger 1964), but, because we are interested in identification, this issue does not concern us here.

Because $\text{var}(x)$ enters the denominator in both (2) and (5) we can derive the relationship between the latent linear model coefficient, β^* , and the linear probability model coefficient, β^{LPM} :

$$\beta^{LPM} = \beta^* \frac{\text{cov}(x, y)}{\text{cov}(x, y^*)}. \quad (6)$$

From (6) we see that the LPM coefficient identifies the “true” LLM coefficient up to a ratio of covariances, $\frac{\text{cov}(x, y)}{\text{cov}(x, y^*)}$. Because the denominator in the covariance ratio equals

$$\text{cov}(x, y^*) = \beta^* \text{var}(x) = \beta^* \sigma_x^2 \quad (7)$$

and because the numerator equals

$$\begin{aligned} \text{cov}(y, x) &= E(xy) - E(x)E(y) = E(xE(y|x)) - E(x)E(y) \\ &= \int xF\left(\frac{\alpha + \beta^* x}{\sigma}\right)g(x)\partial x - E(x)\int F\left(\frac{\alpha + \beta^* x}{\sigma}\right)g(x)\partial x \end{aligned} \quad (8)$$

where σ is a scale parameter equal to the standard deviation of u in (1) and $g(x)$ is the density function for x , we have that

$$\beta^{LPM} = \frac{\int (x - E(x))F\left(\frac{\alpha + \beta^* x}{\sigma}\right)g(x)\partial x}{\sigma_x^2}. \quad (9)$$

The expression in (9) suggests that the LPM coefficient is a complex transformation of the LLM coefficient. However, in the Appendix we show that the sign of the LPM coefficient always equals that of the LLM coefficient, meaning that the following relations hold:

$$\begin{aligned} \beta^* = 0 &\Leftrightarrow \beta^{LPM} = 0 \\ \beta^* > 0 &\Leftrightarrow \beta^{LPM} > 0 \\ \beta^* < 0 &\Leftrightarrow \beta^{LPM} < 0. \end{aligned} \quad (10)$$

As we return to in the concluding section, the result in (10) is useful for understanding what we can compare when using the LPM for comparisons between groups.

To better understand the complex expression in (9), we evaluate it using a Taylor-expansion around small β^* . To illustrate that the issue exists for larger β^* we later modify this assumption in a Monte Carlo study. Using a first-order Taylor expansion of

$$\int (x - E(x)) F\left(\frac{\alpha + \beta^* x}{\sigma}\right) g(x) dx \quad (11)$$

around $\beta^* = 0$ we obtain the following approximation:

$$\beta^* \frac{\sigma_x^2}{\sigma} f\left(\frac{\alpha}{\sigma}\right). \quad (12)$$

From the approximation in (12) we can further approximate the relationship between β^{LPM} and β^* as

$$\beta^{LPM} \approx f\left(\frac{\alpha}{\sigma}\right) \frac{\beta^*}{\sigma}. \quad (13)$$

The expression in (13) reveals that the LPM coefficient recovers the LLM coefficient, first, up to scale, which is the standardising scale factor and, second, up to a transformation of the truncation of the outcome variable. The result in (13) shows that LPM coefficient approximates the marginal effect implied by a nonlinear probability model, as this is usually defined (Wooldridge 2002). Furthermore, the result in (13) can be considered a generalisation of that reported for the logit model by Mare (1981) more than three decades ago.

To further investigate the relationship between β^* and β^{LPM} we evaluate (9) using a second-order Taylor expansion. The Appendix gives the derivation. We find that

$$\beta^{LPM} \approx f\left(\frac{\alpha}{\sigma}\right) \left(\frac{\beta^*}{\sigma}\right) + (\beta^*)^2 f'\left(\frac{\alpha}{\sigma}\right) \left(\frac{E(x^3) - E(x^2)\gamma}{\sigma^2 \sigma_x^2}\right) \quad (14)$$

where γ is the mean of x . While the approximation in (14) is difficult to interpret it shows that the relationship between β^* and β^{LPM} depends on the third moment of x , i.e. it is sensitive to the shape of the distribution of x .

In sum, the LPM coefficient depends on the following four components: the LLM coefficient β^* , the scale parameter σ , the truncation of the outcome α and the distributional shape of x .

4. Group differences in coefficients

The nonlinear relationship between the LPM and LLM coefficients begets the question of what group differences in LPM coefficients identify in terms of the corresponding group differences in LLM coefficients. To answer this question, we first make a simple exposition in terms of comparisons of effects between two groups, then turn to a generalisation in terms derivatives and end with a Monte Carlo study to study cases for which our Taylor-expansions do not hold.

3.1. Difference in coefficients between two groups

To clarify what group comparisons of LPM coefficients identify in terms of the “true” coefficients of the LLM we use the Taylor approximation in (13). Given (13) we can state the difference in LPM coefficients across groups A and B as

$$\beta_A^{LPM} - \beta_B^{LPM} = \beta_A^* \frac{f\left(\frac{\alpha_A}{\sigma_A}\right)}{\sigma_A} - \beta_B^* \frac{f\left(\frac{\alpha_B}{\sigma_B}\right)}{\sigma_B} \quad (15)$$

which is equivalent to the interaction term in a LPM on the pooled sample. The equality in (15) suggests, as we would have expected, that group differences in LPM coefficients need not only reflect the corresponding differences LLM coefficients, but also reflect differences in the truncation of the outcome, α_A vs. α_B , and in the scale parameters, σ_A vs. σ_B . Moreover, if we use the second-order Taylor approximation in (14) as in (15), then it is straightforward to show that group differences in LPM coefficients are also sensitive to group differences in the shape of the distribution of x , a point we return to in a Monte Carlo study.

4.2. Monte Carlo studies

4.2.1. Truncation and scaling differences

To illustrate how group differences in LPM coefficients are affected by group differences in truncation and scaling we used Monte Carlo simulations to study how often the LPM rejected the true null hypothesis of no interaction effect (Type I error). In all scenarios, the error term in the LLM is normal, x is standard normal, we have two groups A and B which is generated from a binary variable constructed from the dichotomisation of a continuous underlying variable, z^* , which follows a normal distribution. An observation belongs to group A if z^* is smaller than or equal to zero, and z equals one otherwise. In all scenarios the coefficients in the LLM are constant across groups, $\beta_A = \beta_B = 1$, and x and z^* has a 0.33 correlation. All LPMs are estimated using OLS and the standard errors are estimated using the Huber-White sandwich estimator.

Table 1 report the number of times out of 10,000 the LPM rejected the true null hypothesis of no interaction effect. Scenario A acts as baseline, with equal scale parameters and equal alphas (see notes in Table 1 for definitions). For all sample sizes, the LPM rejects the true null around the expected 500 times, correctly reflecting the interaction term in the LLM. In Scenarios B1-B2 we study the performance of the LPM when the groups differ in α 's, i.e. in the mean of the latent propensity y^* for the average x person, holding constant the scale parameters. In B1, in which alphas differ by a small amount, we find that the LPM rejects the true null more than 500 times for all sample sizes, but dramatically exceeds 500 times for large samples sizes. Scenarios C1-C2 holds constant the alphas and varies the scale parameters, i.e. the extent to which the latent residual variation differs across groups. In C1, the scale parameters differ only marginally with a factor of 1.1 but for all sample sizes the true null is rejected more than the expected 500 times. For $N = 5,000$ the rejection rate is 30 percent. In

C2, the scale parameters differ with a factor of 1.5 and the rejection rate pattern is even stronger: for all sample sizes, the rejection rate is more than the expected 5 percent.

These results corroborate our analytical results. The LPM is sensitive to differences in truncation and scale parameters across groups, in particular in large samples in which the LPM rejects the true null more than the expected 500 times.⁵ The simulation study thus indicates that inference on the group differences can be erroneous, and one is likely to commit a Type I error (rejecting a true null). Furthermore, although the Monte Carlo study we report here is far from exhaustive it is easy to construct examples in which using the LPM for inference about the LLM leads to Type II errors (accepting a false null of no group differences in the LLM). We turn to such case in the Example section.

-- TABLE 1 HERE --

4.3.2. Differences in the distributional shape of x

As our second-order Taylor expansion suggested differences in the shape of the distributions of x across groups might also affect interaction terms in the LPM. To study this situation we conducted a Monte Carlo study, using Scenario A in Table 1 as the baseline design (i.e. we assume that all latent parameters are equal). To examine whether group differences in the distributional shape of x affect the interaction term in the LPM, we transform x to a log-normal distribution and vary the correlation between x and z^* . This nonlinear specification allows the distributions of x to differ as we vary the correlation between x and z^* .

⁵ Note that this result is consistent with the example in Allison (1999), where a significant effect in a homoscedastic logit turns into a zero effect in heteroscedastic logit model.

Table 2 presents the results. In the first column in which x and z^* are independent we find the expected 5 percent rejection rate in large samples ($N=5,000$), but not in small samples ($N=100$). This finding suggests that, in large samples and when distribution of x is identical across groups, the interaction term in LPM correctly reflects the corresponding interaction terms in the LLM. However, as we increase the correlation between x and z^* and consequently let the distributional shape of x differ across groups the rejection rates exceed the expected 5 percent, in particular in large samples. These results confirm the analytical result that group differences in the LPM can reflect differences in the distribution of x across groups.

-- TABLE 2 HERE --

5. Examples

To illustrate some of the issues encountered in using LPMs in group comparisons we use data from the U.S. General Social Survey (GSS) 1988-2010 to examine historical trends in, first, the black and white gap in educational attainment and, second, the economic returns to education in the US. In the first example, we take years of education as our continuous outcome variable, reflecting a latent propensity to educate oneself. As our binary realisation of this variable we use 16 years as our cut point; the equivalent of a four-year college degree in the US. In the second example we take log individual income as a continuous outcome variable and the birth cohort specific median of the income variable as our cut-point for the binary realisation.

5.1 Trends in black and white achievement gap

In our first example, we investigate whether the black and white gap in educational attainment has changed across two cohorts born before and after the Second World War. We use a sample

of blacks and whites, aged 25 to 65, born in one of two birth cohorts, 1925-1944 or 1945-1959 and with valid information on the educational attainment variable, yielding a total sample of 25,308 respondents. In the models blacks take on the value 1 and whites take on the value 0, meaning that we examine how blacks fare relative to whites.

Table 3 reports the results. The model for years of education (LLM) shows a marked and statistically significant decline in the black and white gap between those born 1925-1944 and those born 1945-1959. The gap dropped from 1.4 to 0.8 years of education; a 43 percent reduction. Compare this result with that obtained from the second model on whether or not the respondent completed 16 years of education (LPM). The gap, 13 percentage points, is virtually the same in the 1925-1944 and 1945-1959 cohorts, and the small change (-0.005) is far from statistically significant on conventional levels. In this case the results markedly differ between the two models, suggesting that the LPM does not capture the true heterogeneity in effects and thereby committing a Type I error (rejecting a true null).

-- TABLE 3 HERE --

To better understand why the models' results differ, in Table 4 we report three distributional characteristics for each cohort. Variations in these characteristics between cohorts can help to explain the results. We find that the conditional standard deviation of years of education given race – i.e. the scale parameter (σ) – declined from 3.1 in the pre-War cohort to 2.8 in the post-War cohort, and that the average years of education completed (α) rose from 12.5 to 13.6. These differences between the two cohorts suggests that either the decrease in the scale parameter or the increase in average years of education or both mask the true decline in the black and white gap over the 20th century US.

-- TABLE 4 HERE --

5.2 Trends in the returns to education

In our second example, we examine trends in the returns to education. Similar to the first analysis we use a sample of blacks and whites, aged 25 to 65, and born between either 1925-1944 or 1945-1959. We further restrict the sample to respondents with valid information on age, gender and individual income in constant dollars. The total sample comprises 17,277 respondents. In the linear models, we use log to individual income as our outcome variable, and for the LPMs we cut the income distribution at the respective medians of each cohort. As a robustness check, we also report the effects on the median of the log to individual income using quantile regression. This check allows us to ascertain that our LPM does not simply gauge the effects on the median but rather reflects – correctly or incorrectly – the configuration of the parameters in the linear model (LLM).

In Table 5 we report the results. The coefficients reported in the linear model (LLM) are elasticities, measuring the percentage change in income for a one year change in education net of age, race and gender. We find that the returns to education in these models change very little across the birth cohorts; a result corroborated by the statistically insignificant cohort difference in coefficients. Thus according to this model the returns to a year of education are roughly 10 percent in both cohorts. By way of contrast in the linear probability model on the dichotomised income variable at the median (LPM), it appears that the returns to education significantly increased from the 1925-1944 cohort to the 1945-1959 cohort. The increase is 0.5 percentage points, amounting to about a 10 percent increase. We consequently find evidence of a Type I error (rejecting a true null) in this example.

-- TABLE 5 HERE --

To check whether the discrepancy in results between the two models is an artefact of distributional effects, in the final column in Table 5 we report the results for a quantile regression modelling the median of log to individual income. Similar to the linear model (LLM) the changes in the returns to education are constant between the two cohorts. Thus the significant and marked change in returns found in the LPM cannot be explained by potential difference in the distributional effects on the log income margin.

To further examine the potential reasons for the difference in results for the two models, in Table 6 we report distributional characteristics that might explain the results. The scale parameters (σ) and the skewness of the years of education variable appear to change little between the cohorts, placing the reason for the difference in results in the truncation (α). We find that the difference between the mean log income (α) and the median (the cut-point) decreases across the cohorts. In constant dollars (fourth column in Table 6), we see a marked drop of roughly 1,000 dollars between cohorts, amounting to a 20 percent reduction. This discrepancy likely explains the difference in results generated by the LLM and LPM.

6. Conclusion

Analysing group differences in effects is central to much inquiry in social science research. In this article we formally show the challenges when using linear probability models for comparing coefficients across groups. We find that comparisons of LPM coefficients suffer not only from the scale identification problems pertaining to the logit or probit, but also from problems pertaining to outcome truncation and distributional shape of the predictor. As a consequence magnitudes and directions of group differences in LPM coefficients are generally uninformative about group differences in the parameters assumed to have generated the data.

Nevertheless, our paper shows that the sign of the LPM coefficient always reflects the sign of the LLM coefficient. One important implication of this result for applied research is that researchers should compare only the direction of the LPM coefficient across groups. For example, if the effect is found to be nil for one group, but positive for another, this result might provide valuable information on group differences in whether or not a variable affects the outcome of interest. Thus in situations where coefficient signs differ between groups the LPM is informative – although in a restricted way – about group difference in effects. However, in cases where researchers need to compare the size of the effect of a variable across groups one need continuous outcomes. Hence, whenever the researcher is in control of the outcome variable, either when sampling the data or trough formulating a research question, a continuous outcome should be strongly preferred.

References

- Ai, C. and Norton, E. C. (2003) Interaction terms in logit and probit models. *Economics Letters*, 80, 123-129.
- Allison, P. D. (1999) Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods & Research*, 28, 186-208.
- Amemiya, T. (1981) Qualitative Response Models: A Survey. *Journal of Economic Literature*, 19, 1483-1536.
- Anwar, S., P. Bayer and R. Hjalmarrsson (2012) The Impact of Jury Race in Criminal Trials, *Quarterly journal of Economics*, 127, 1017-1055.
- Athey, S. and Imbens, G. W. (2006) Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, 74, 2006.
- Breen, R., Holm, A and K. B. Karlson (2014). Correlations and non-linear probability models. Forthcoming, *Sociological Methods and Research*.
- Goldberger, A. S. (1964) *Econometric Theory*. New York: John Wiley and Sons.
- Fisman, R., S. S. Iyengar, E. Kamenica and I. Simonson (2006) Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment. *Quarterly Journal of Economics*, 673-607.

Greene, W. H. (2010) Testing hypotheses about interaction terms in non-linear models. *Economics Letters*, 107, 291-296.

Greene, W. H. (2011) *Econometric Analysis*, 7th Edition. Upper Saddle River, NJ: Prentice Hall.

Gould, E. D. and E. F. Klor (2010) Does Terrorism Work? *Quarterly Journal of Economics*, 1459-1510.

Karaca-Mandic, P., Norton, E. C. and Dowd, B. (2011) Interaction Terms in Nonlinear Models. *Health Services Research*, 47, 255-274.

Lewbel, A., Y. Dong, and T. T. Yang (2012) Comparing Features of Convenient Estimators for Binary Choice Models With Endogenous Regressors,” *Canadian Journal of Economics*, 45, 809-829.

Long, J.S. (2009) Group comparisons in logit and probit using predicted probabilities. *Unpublished working paper (June 25, 2009)*.

Mare, R. D. (1981) Change and Stability in Educational Stratification. *American Sociological Review*, 46, 72-87.

Mare, R. D. (2006) Response: Statistical Models of Educational Stratification: Hauser and Andrew's Model for School Transitions. *Sociological Methodology*, 36, 27-37.

Meghir, C and Palme, M. (2005) Educational Reform, Ability, and Family Background, *American Economic Review*, 95, 414-424.

Norton, E. C., Wang, H. and Ai, C. (2004) Computing interaction effects and standard errors in logit and probit models. *The Stata Journal*, 4, 154-167.

Allison, P. (1999) *Logistic Regression Using SAS: Theory and Application*. Cary, NC: SAS Institute.

Swait, J. and Louviere, J. (1993) The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research*, 30, 305-314.

Williams, R. (2009) Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients Across Groups. *Sociological Methods & Research*, 37, 531-559.

Vigna, S. D., J. A. List and U. Malmendier (2012) Testing for Altruism and Social Pressure in Charitable Giving, *Quarterly Journal of Economics*, 27, 1-56.

Wooldridge, J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*, MIT press.

Xie, Y. (2011) Values and limitations of statistical models. *Research in Stratification and Social Mobility*, 29, 343-349.

Appendix

Identical sign of LPM and LLM coefficients

We want to show (10), i.e.

$$\begin{aligned}\beta^* = 0 &\Leftrightarrow \beta^{LPM} = 0 \\ \beta^* > 0 &\Leftrightarrow \beta^{LPM} > 0 \\ \beta^* < 0 &\Leftrightarrow \beta^{LPM} < 0.\end{aligned}$$

From (11) we have that

$$\int (x - E(x)) F\left(\frac{\alpha + \beta^* x}{\sigma}\right) g(x) \partial x$$

We show that $\beta^* > 0 \Rightarrow \beta^{LPM} > 0$. The situation in which $\beta^* < 0 \Rightarrow \beta^{LPM} < 0$ follows in an equivalent way.

Assume that $\beta^* > 0$, which implies that $F(\cdot)$ is an increasing function of x .

Rewrite (11) as

$$\begin{aligned}\int (x - E(x)) F\left(\frac{\alpha + \beta^* x}{\sigma}\right) g(x) \partial x = \\ \int_{-\infty}^{E(x)} (x - E(x)) F\left(\frac{\alpha + \beta^* x}{\sigma}\right) g(x) \partial x + \int_{E(x)}^{\infty} (x - E(x)) F\left(\frac{\alpha + \beta^* x}{\sigma}\right) g(x) \partial x\end{aligned}$$

from which we obtain

$$\begin{aligned}\int (x - E(x)) F\left(\frac{\alpha + \beta^* x}{\sigma}\right) g(x) \partial x > \\ \int_{-\infty}^{E(x)} (x - E(x)) F\left(\frac{\alpha + \beta^* E(x)}{\sigma}\right) g(x) \partial x + \int_{E(x)}^{\infty} (x - E(x)) F\left(\frac{\alpha + \beta^* E(x)}{\sigma}\right) g(x) \partial x\end{aligned}$$

We can then evaluate each of the terms. For the first term we have that $(x - E(x)) < 0$ and

that $F\left(\frac{\alpha + \beta^* x}{\sigma}\right) < F\left(\frac{\alpha + \beta^* E(x)}{\sigma}\right)$ for $x < E(x)$. This further implies that

$$\int_{-\infty}^{E(x)} (x - E(x)) F\left(\frac{\alpha + \beta^* E(x)}{\sigma}\right) g(x) \partial x < \int_{-\infty}^{E(x)} (x - E(x)) F\left(\frac{\alpha + \beta^* x}{\sigma}\right) g(x) \partial x.$$

For the second term, we have that $(x - E(x)) > 0$ and that $F\left(\frac{\alpha + \beta^* x}{\sigma}\right) < F\left(\frac{\alpha + \beta^* E(x)}{\sigma}\right)$

for $x > E(x)$ and that:

$$\int_{E(x)}^{\infty} (x - E(x)) F\left(\frac{\alpha + \beta^* E(x)}{\sigma}\right) g(x) dx < \int_{E(x)}^{\infty} (x - E(x)) F\left(\frac{\alpha + \beta^* x}{\sigma}\right) g(x) dx .$$

It then follows that

$$\int_{-\infty}^{\infty} (x - E(x)) F\left(\frac{\alpha + \beta^* x}{\sigma}\right) g(x) dx > \int_{-\infty}^{\infty} (x - E(x)) F\left(\frac{\alpha + \beta^* E(x)}{\sigma}\right) g(x) dx = 0$$

Next we show the case when $\beta^* = 0 \Rightarrow \beta^{LPM} = 0$. This follows straight forward from

$$\int (x - E(x)) F\left(\frac{\alpha}{\sigma}\right) g(x) dx = 0 .$$

Now we have established that

$$\begin{aligned} \beta^* > 0 &\Rightarrow \beta^{LPM} > 0 \\ \beta^* = 0 &\Rightarrow \beta^{LPM} = 0 \\ \beta^* < 0 &\Rightarrow \beta^{LMP} < 0 \end{aligned}$$

To show that $\beta^{LMP} > 0 \Rightarrow \beta^* > 0$ we can instead show that $\beta^* \leq 0 \Rightarrow \beta^{LPM} \leq 0$,⁶ If $\beta^* \leq 0$

then either $\beta^* = 0$ or $\beta^* < 0$. If $\beta^* = 0$ we know from equation (10) that $\beta^{LPM} = 0$.

Similarly, we have that $\beta^* < 0$ leads to $\beta^{LPM} < 0$. Then we have that $\beta^{LPM} \leq 0$. In a similar

way we can show the $\beta^{LPM} = 0 \Rightarrow \beta^* = 0$ and $\beta^{LPM} < 0 \Rightarrow \beta^* < 0$.

We have thus shown:

$$\begin{aligned} \beta^* > 0 &\Leftrightarrow \beta^{LPM} > 0 \\ \beta^* = 0 &\Leftrightarrow \beta^{LPM} = 0 \\ \beta^* < 0 &\Leftrightarrow \beta^{LMP} < 0 \end{aligned}$$

⁶ We are here using that showing that $A \Rightarrow B$ is equivalent to show that $\neg B \Rightarrow \neg A$.

Second-order Taylor approximation

With one predictor variable, x , with mean γ , we derive $\text{cov}(y,x)$:

$$\begin{aligned}\text{cov}(y, x) &= E(xy) - E(x)E(y) = E(x | E(y | x)) - E(x)E(y) \\ &= \int xF\left(\frac{\alpha + \beta^* x}{\sigma}\right) f(x) dx - \gamma \int F\left(\frac{\alpha + \beta^* x}{\sigma}\right) g(x) dx.\end{aligned}\tag{A1}$$

We approximate (A1) by a second-order Taylor approximation around $\beta^* = 0$:

$$\begin{aligned}\text{cov}(y, x) &\approx \int xF\left(\frac{\alpha}{\sigma}\right) + x\beta^* f\left(\frac{\alpha}{\sigma}\right)\frac{x}{\sigma} + (\beta^*)^2 x f'\left(\frac{\alpha}{\sigma}\right)\left(\frac{x}{\sigma}\right)^2 g(x) dx \\ &\quad - \gamma \int \left(F\left(\frac{\alpha}{\sigma}\right) + \beta^* f\left(\frac{\alpha}{\sigma}\right)\frac{x}{\sigma} + (\beta^*)^2 f'\left(\frac{\alpha}{\sigma}\right)\frac{x^2}{\sigma^2} \right) g(x) dx \\ &= F\left(\frac{\alpha}{\sigma}\right)\gamma + \beta^* f\left(\frac{\alpha}{\sigma}\right)\left(\frac{E(x^2)}{\sigma}\right) + (\beta^*)^2 f'\left(\frac{\alpha}{\sigma}\right)\left(\frac{E(x^3)}{\sigma^2}\right) \\ &\quad - \left(\lambda F\left(\frac{\alpha}{\sigma}\right) + \beta^* f\left(\frac{\alpha}{\sigma}\right)\frac{\gamma^2}{\sigma} + (\beta^*)^2 f'\left(\frac{\alpha}{\sigma}\right)\frac{E(x^2)\gamma}{\sigma^2} \right) \\ &= \beta^* f\left(\frac{\alpha}{\sigma}\right)\left(\frac{E(x^2) - \gamma^2}{\sigma}\right) + (\beta^*)^2 f'\left(\frac{\alpha}{\sigma}\right)\left(\frac{E(x^3) - E(x^2)\gamma}{\sigma^2}\right) \\ &= f\left(\frac{\alpha}{\sigma}\right)\left(\frac{\beta^* \sigma^2_x}{\sigma}\right) + (\beta^*)^2 f'\left(\frac{\alpha}{\sigma}\right)\left(\frac{E(x^3) - E(x^2)\gamma}{\sigma^2}\right),\end{aligned}$$

yielding (13).

TABLES

Table 1. Monte Carlo simulation of interaction terms in LPM. No of times rejecting true two-sided null (Type-I error) out of 10,000. Normal x and z^* .

	$N = 100$	$N = 1,000$	$N = 5,000$
A	540	534	504
B1	588	1,049	3,214
B2	2,738	9,853	10,000
C1	573	1,040	3,067
C2	1,336	7,667	10,000

Notes: 500 equals conventional 5-percent significance level. z^* denotes a latent version of binary z .

Scenarios:

A: $\sigma_A = \sigma_B = 1$; $\alpha_A = \alpha_B = 0$; $\beta_A^* = \beta_B^* = 1$; $\lambda_A \neq \lambda_B$; $R(x, z^*) = 0.33$

B: $\sigma_A = \sigma_B = 1$; $\alpha_A \neq \alpha_B$; $\beta_A^* = \beta_B^* = 1$; $\lambda_A \neq \lambda_B$; $R(x, z^*) = 0.33$

B1: $\alpha_A = 0.25$; $\alpha_B = 0$

B2: $\alpha_A = 1$; $\alpha_B = 0$

C: $\sigma_A = \sigma_B = 1$; $\alpha_A \neq \alpha_B$; $\beta_A^* = \beta_B^* = 1$; $\lambda_A \neq \lambda_B$; $R(x, z^*) = 0.33$

C1: $\sigma_A = 1.1$; $\sigma_B = 1$

C2: $\sigma_A = 1.5$; $\sigma_B = 1$

Table 2. Monte Carlo simulation of interaction terms in LPM. No of times rejecting true two-sided null (Type-I error) out of 10,000. Lognormal x and normal z^* .

	$N = 100$	$N = 1,000$	$N = 5,000$
$R(x, z^*) = 0.00$	863	588	511
$R(x, z^*) = 0.33$	910	888	1,919
$R(x, z^*) = 0.66$	948	1,657	5,326

Notes: 500 equals conventional 5-percent significance level. $R(x, z^*)$ denotes the correlation before x is

transformed. z^* denotes a latent version of binary z . Scenario: $\sigma_{z=1} = \sigma_{z=0} = 1$; $\alpha_{z=1} = \alpha_{z=0} = 0$; $\beta_{z=1}^* = \beta_{z=0}^* = 1$.

Table 3. Black and white gaps in educational attainment. Three birth cohorts. Coefficients express how much better or worse blacks fare compared to whites.

<i>Cohort</i>	LLM Years of education		LPM > 16 years of education	
	Coef.	t	Coef.	T
1925-1944	-1.387	-13.94	-0.126	-14.17
1945-1959	-0.796	-11.80	-0.131	-12.79
Difference	0.592	4.92	-0.005	-0.34

Table 4. Scale parameter (conditional variance in years of education given race) and truncation (average years of education) in two cohorts.

	Scale parameter σ	Truncation α
1925-1944	3.112	12.455
1945-1959	2.758	13.574

Table 5. The effect of years of education on log to individual income (LLM), on whether or not income exceeds the median in the cohort (LPM) and on median log individual income in cohort (Quantile regression).

	LLM ln(income)		LPM > Median in cohort		Quantile regression Median	
	Coef.	t	Coef.	t	Coef.	t
1925-1944	0.0996	27.06	0.0483	27.27	0.0939	29.70
1945-1959	0.1038	28.43	0.0540	31.91	0.1006	35.77
Difference	0.0042	0.82	0.0057	2.33	0.0067	1.59

Note: All estimates adjusted for age, gender, and race.

Table 6. Scale parameter (conditional variance log income given education, race, gender and age), truncation (average years of education, median and their difference) and skewness of education variable in two cohorts.

	Scale parameter Σ	Truncation α	Truncation Median	Truncation: α -Median in USD	Skewness: Education
1925-1944	0.8724	10.1048	10.2997	-5,264	-0.0421
1945-1959	0.8734	10.1068	10.2635	-4,159	-0.0203